METRICS

Thomas Fitzmaurice
Six Sigma Master Black Belt

4/13/2006

Objective

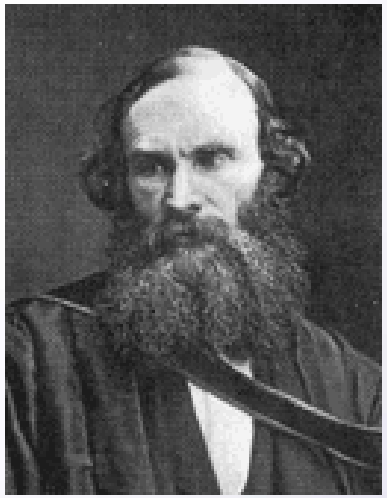
Understand the elements a reliable and accurate metric.

Agenda

- Why do We Need Metrics
- Characteristics of Good Business Metrics
  - Purpose
  - Validity
  - Reliability
  - Accuracy
  - Sensitivity
- Example of Metric Usage

Why Do We Need Metrics?

- "To measure is to know."
- "If you can not measure it, you can not improve it"
- “The first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it “



William Thompson
Lord Kelvin
(1824-1907)

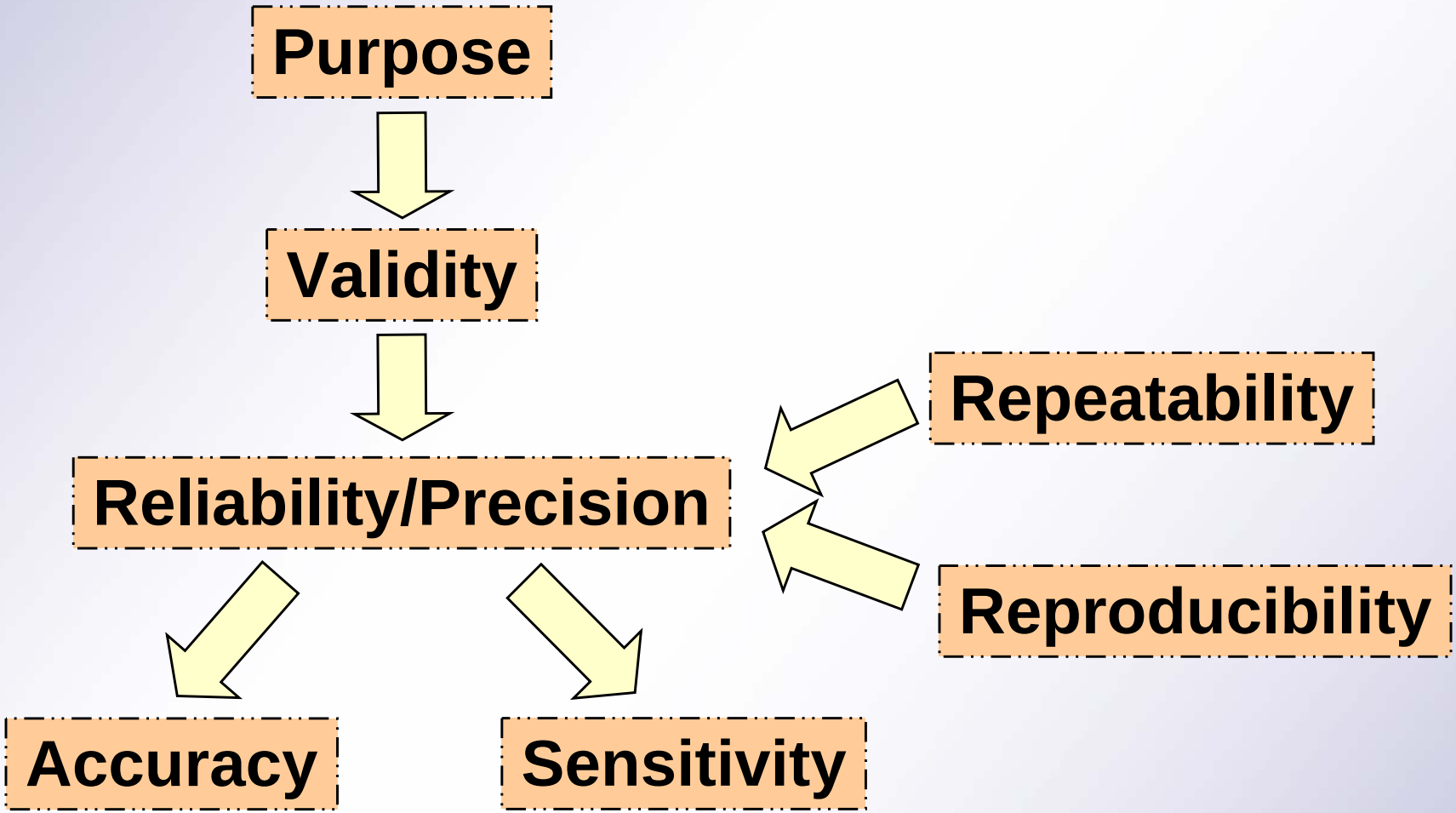
“I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you can not measure it, when you can not express it in numbers, your knowledge is of a meagre (sic) and unsatisfactory kind"

Using Metrics

- Measurements and metrics are created to satisfy some type of knowledge need.
- Ultimately, metrics are likely to be used to help make decisions in the organization. It is important to understand how metrics are perceived and used during the decision making process.
- To clearly define their purpose we must know the who, what, where, when, why, and how for each of the decisions in which they are used.
- In the absence of such understanding, there will always be doubts about the function and utility of many business metrics.

Metrics Help Us Balance Risk and Cost

Hierarchy of Metric Characteristics



Individuals in an organization will not use metrics that they do not believe in. In order to be believable, metrics must have these characteristics.

Purpose

Every metric must have purpose:

- Typically, metrics provide business value by providing information, facts or knowledge to help make business decisions.

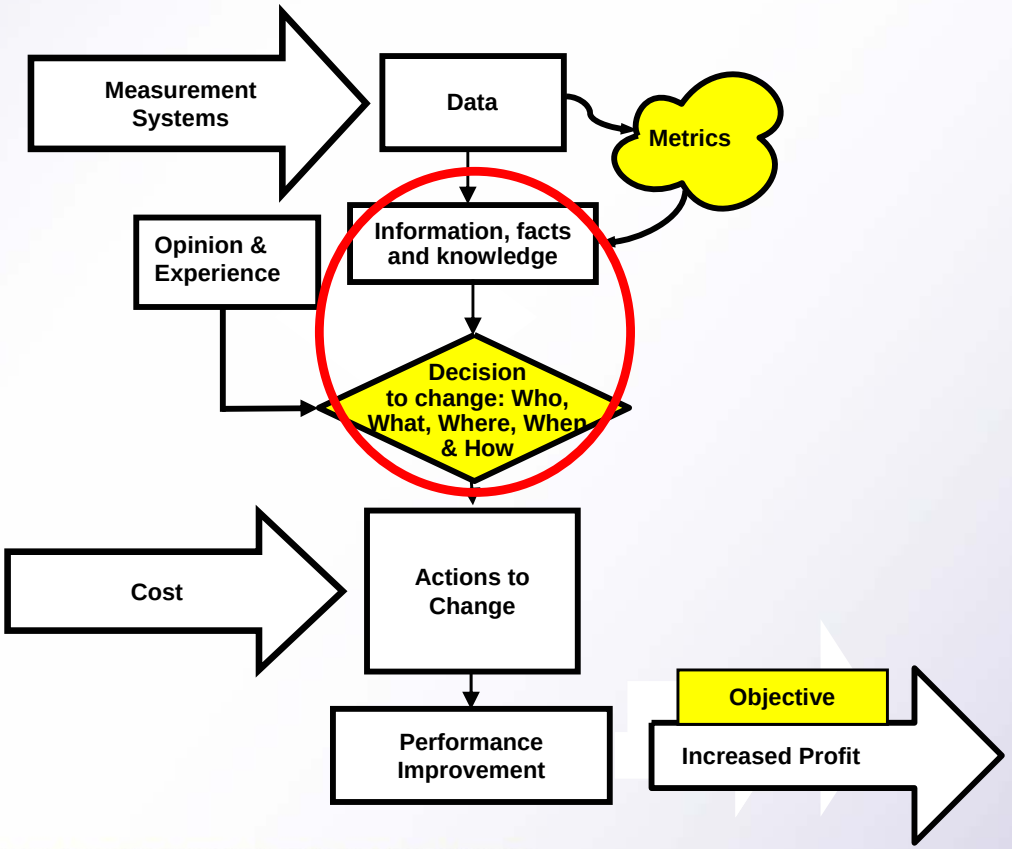
- To clearly define the purpose we must:
  - Understand what is important to our customer, our business and our company (business requirements).
  - Know the who, what, where, when, why, and how for each of the decisions in which the metric will be used.
  - Ask the question: “What information do I need to know to make the decision?”
  - In the absence of such understanding, there will always be doubts about the function and utility of many business metrics.

Too often, organizations build complex data collection and information management systems without truly considering whether the data collected and metrics calculated actually benefit the organization.

Purpose (Cont.)

Establish a purpose—"What do I need to know!"

- Ideally, the defined purpose should relate to a true need of the customer, business and/or company. Since the implied function of all business metrics is to be an input in some type of decision, focus on decision points in the process to help establish need/purpose



Purpose (Cont.)

Some questions to ask/consider include:

- How will the metric provide business value?
  - What business requirement/strategy is being addressed?
  - What is important to the customer, business and/or company?
- What decisions will this metric be used for?
  - Exactly, what information is needed?
  - How important is the metric/information to this decision?
  - What is the decision criteria (target, limits, specs.)?
  - Are there other metrics involved in the decision?
  - What are options or alternatives? How many different outcomes of the decision are possible?
- Who are the decision makers?
  - How are they motivated (rewarded or punished) for their decision performance?

Validity

Validity can be classified into four lower level categories:

- Construct Validity: Metric meets the purpose for which it was developed (decision making).
  - There is understanding of who makes the decisions
  - There is clear understanding of the decision criteria
  - Consequences of a bad decision are understood

- Content Validity: Metric measures the right thing (property)

- Face Validity: Measures what you say it is measuring.
  - Subjective and consequently the weakest form of validity (e.g. survey data).
  - But, provides a good “sanity check”

- Statistical Validity: Statistical data where the appropriate tests are applied and there is full compliance to all the required assumptions (e.g. type of data, normality, independence,…).

Validity (Cont.)

- If the metric has validity:
  - Good business decisions are possible
  - The risks and consequences of bad decisions are understood
- If the metric is not valid:
  - The metric will either be ignored or possibly misused
  - Bad decisions can be made
  - Evaluate and reconsider:
    - The property of the data being measured
    - The information being provided by the metric
    - The decision criteria being used in the metric

Validity (Cont.)
Methods of Measurement

There are two general ways that we can collect data:

1. Direct Measurement
   - We physically measure one or more characteristics of a system or entity. Often these measurements are physical properties; dimensions, quantity or capacity. (e.g. size, time, cost, count, defects, temperature, …)

2. Indirect Measurement
   - We measure one characteristic as an indicator or proxy of another characteristic. Often these measurements are of human characteristics. An example is, measuring the quality of engineering deliverables by counting the number of Field Change Requests received

   - Indirect measurements can lead to significant validity problems. The relationship between the indicator variable and the target variable must be well understood

Reliability/Precision

A reliable metric is one that returns the same value of a given property every time a measurement is made. Reliable metrics are independent of who estimates them or when the estimate is made. Reliability can be partitioned into two components:

- **Repeatability**
  - A single person or measurement system consistently returns the same value when measuring an identical property on the same part or item
- **Reproducibility**
  - Multiple people or measurement systems consistently return the same value when measuring an identical property on the same part or item using the same device

Reliability/Precision (Cont.)

- Reliable metrics encourage:
  - Belief
  - Confidence
  - Trust
- Metrics that lack reliability result in:
  - Disbelief
  - Distrust
  - And, typically bad decisions
- A Gage Study is used to determine the reliability of a metric
  - (not discussed in this presentation)

Accuracy

- Accuracy is the level of exactness or correctness, consistent with a standard, rule or model.

- An accurate measurement is one that matches the true value of a property. When a metric or measurement system consistently over or under estimates the value of a property, it is said to be “inaccurate”.

- Accuracy can be assessed in several ways:
  - Measurement of a known standard
  - Comparison with another known measurement method
  - Prediction of a theoretical value

- What happens if we don’t have standards, comparisons, or theories?

- If your data is not accurate, then you may not be able to make good business decisions.
  - If, for example, your metric consistently overestimates a cost, you may choose not to implement a beneficial change, when, in fact, the return on investment justifies it.

Accuracy (Cont.)

- Every Metric is a Statistic (it is an estimate based on a summary of data)
- Every statistic has a sampling theory!
- Use the sampling theory to understand how the quantity of data that you have effects the amount of uncertainty in the metric.
  - More data = higher confidence
- For example: A set of results of a Gallop Poll may tell us that the uncertainty of a certain survey result is approximately ± 4%.
  - This uncertainty tells us that a survey result of 90% could produce an actual value as high as 94% or as low as 86%—strictly through luck!

Sensitivity

Measurement sensitivity refers to the ability of a metric to detect change. Sensitivity can be effected by many factors including:

1. Measurement resolution or unit of measure
   - Cycle times measured in days will not be sensitive to changes of minutes or hours
   - Resolution should consider practicality. Is it practical to measure the cycle time of a routine document review in seconds? minutes? hours? 8 hour shifts? days?

2. Type of statistic used
   - The median, as a statistic, is not very sensitive to change. The mean, on the other hand, is extremely sensitive to change. The mathematical form of your metrics calculations can have a large impact on sensitivity.

3. Type of data used
   - Metrics based on counts are generally less sensitive to change than those based on actual measurements.

Sensitivity (Cont.)

- If your data is not sensitive enough, you may not react to meaningful changes in your process (low cost high risk)
- If your data is too sensitive, you may over-react and implement controls or procedures that are really not necessary, spending unnecessary money
- You must find the right balance:
  - Ask the question: “What is practically significant?”
    - What is important to the customer, business and/or company.
    - Where is the baseline performance?
    - What was established as the target improvement?
    - What is the Gap? The smaller the gap, the smaller the resolution!
    - Example: If the performance gap is 38 days, measuring in hours is of no practical value. Resolution should be days

Example of Metric Usage

Scenario

Your customer receives a call from a former disgruntled employee indicating that workers are routinely expected to work when their training qualifications are expired. The employee worked in Nuclear Operations.

You are the manager assigned to handle the issue.

What do you do?

Typical Responses

- Do nothing.
- Do nothing and attack the integrity of the whistle blower.
- Assume it is true and fire all those involved.
- Assume it is true and conduct an operational stand down and emphasize the need to maintain qualified workers.
- Conduct surveys of workers who are currently employed and base your actions off the survey information. (Survey information can be extremely biased and therefore are generally considered unreliable.)

Metric Based Response

- Purpose
  - Understand the validity of the accusation in order to formulate the appropriate response.
  - To understand the validity of the accusation a measure/metric of whether workers are performing work with expired training and if so what is the magnitude of the problem.

- Data needed includes:
  - Training data (completion dates and expiration dates for workforce)
  - Training requirements to perform the scope of work over the last year for the work force.

The Metric needed to make the response decision was; the percentage of workers performing work with expired training in Nuclear Operations.

Data Collection

- Most companies have many processes that collect data. These data sources should be evaluated before another data collection system is developed to support a decision metric (data costs money)

- It is estimated that the NTS has over 600 electronic databases.

The training department audits work execution once a month and compares the work being performed against the electronic database of training records for the entire company (quasi-direct measure).

Does it answer the question I am asking? Is it accurate, precise, and sensitive enough (valid)?

Data Validation

- The training audits depend on three sources of data
  - Training records.
  - Work performed.
  - Training requirements for the work performed.

- All three sources of data must be audited (gauged) for validity.

- The data in the training records database is input by hand. The database was sampled (n=400, electronic records were compared against course sign in sheets on archive). The training database was found to be 99.5% accurate. (Based on the binomial probability distribution the 95% confidence interval was 98.2 to 99.9%)

Data Validation (Cont.)

- Training audited 329 work opportunities in 2004. The work packages and requirements needed to execute the work packages were re-evaluated. The re-evaluation agreed with the original audit 100%.

- The data was determined to be accurate and precise.

What does the data say?

Metric Analysis

- Out of 329 work audits 8 were identified where training requirements were expired (2.4 % of the work audited found workers with expired training). Based on binomial probability density, the 95% confidence interval ranges from 1.1 to 4.1%. We can be 95% confident that the actual percentage of expired training is between 1.1 and 2.4%.

- 172 of the audits were performed in Nuclear Operations. Zero were found with expired training. Based on binomial probability density, the 95% confidence interval ranges from 0.0 to 1.7%.

- No evidence of expired training in Nuclear Operations was found and the metric indicates that if it does occur the percentage is extremely low.

Now what is your response?

Summary

- All measurements need to have business purpose
  - There is value in good information to make good decisions
- Metrics must have:
  - Validity
  - Reliability/Precision
  - Accuracy
  - Sensitivity
- The best way to collect data is with the Direct Method